An Addendum to "How Good is PSK for Peak-Limited Fading Channels in the Low-SNR Regime?"

Wenyi Zhang, *Member, IEEE* [*]

Abstract

A proof is provided of the operational achievability of $R_{\rm rt}$ by the recursive training scheme in [1], for general wide-sense stationary and ergodic Rayleigh fading processes.

1 Introduction

In [1], an issue remaining open is whether the recursive training scheme interpretation [1, Sec. III.B] is operational for fading processes with an infinitely long memory length. As commented in [1], the main difficulty in analysis pertains to the effect of error propagation, which cannot be readily circumvented in order to rigorously establish a channel coding theorem. In this addendum, we provide a proof of the operational achievability of $R_{\rm rt}$ by the recursive training scheme, for general wide-sense stationary and ergodic Rayleigh fading processes, without the m-dependence constraint as further imposed in the partial proof in [1]. The present proof exploits certain techniques in mismatched decoding, and its basic idea differs considerably from that in the previous proof for m-dependent fading processes.

Before proceeding to our proof, we briefly sketch a heuristic reasoning which has often been invoked to argue that the recursive training scheme is operational for channels with memory, and point out why such a heuristic reasoning is insufficient to lead to a mathematically rigorous proof. Consider the recursive training scheme as illustrated in [1, Fig. 1]. The heuristic reasoning is as follows. For general fading processes that possess infinite

---

memory, some residual correlation remains within each parallel sub-channel (PSC) among its K symbols. Due to the ergodicity of the fading process, as we let the interleaving depth L grow to infinity, this correlation asymptotically vanishes and thus each PSC can then be viewed as essentially memoryless. Meanwhile, as we let the coding blocklength K grow to infinity, the channel coding theorem (*e.g.*, [2]) for the L asymptotically memoryless PSCs ensures that each PSC can achieve a rate arbitrarily close to the input-constrained capacity of a true memoryless fading channel, and consequently the arbitrarily reliably decoded symbols can be used as essentially error-free training pilots for the recursive channel estimation/decoding procedure.

Unfortunately, the outlined heuristic reasoning is inherently flawed. It does not properly resolve a tension between channel memory decorrelation and error propagation. The essentially memoryless PSCs are obtained at the cost of indefinitely increasing the interleaving depth L. As L increases, the block error probability of decoding each PSC is required to decrease inversely proportionally with L, in order to prevent catastrophic error propagation, and thus the coding block length K is required to increase correspondingly. However, it is unclear how valid the "essentially memoryless" property of PSCs is as K scales with L asymptotically, because a formal mathematical characterization of such a property would involve a limiting argument that fixes a sufficiently large K and subsequently lets L grow large. At this point, we have observed a logic loop between the scaling of K and L. So even if the heuristic reasoning were indeed correct, it lacks a mathematical confirmation, except for fading processes with finite memory, namely, the so-called m-dependent fading processes.

In our proof in this addendum, we break the logic loop by not requiring the PSCs be essentially memoryless, and hence relieve the tension between channel memory decorrelation and error propagation, since K and L do not need to scale with each other. However, without the essentially memoryless condition, we can no longer ignore the residual correlation within each PSC. Fortunately, the generalized mutual information (GMI) in the theory of mismatched decoding (see, *e.g.*, [3] and references therein) provides us an alternative way of establishing the achievability of information rates (at least) up to the input-constrained capacity of an associated memoryless fading channel, as will be elaborated shortly.

2 Proof

Consider the recursive training scheme as described in [1, Sec. III.B]. We seek to prove that, for an arbitrarily small decoding error rate $\delta > 0$ and an arbitrarily small overall rate loss factor $\lambda > 0$, there exists a code such that the recursive coding scheme achieves the rate $(1 - \lambda) \cdot R_{\mathrm{rt}}$ with an average decoding error probability no greater than δ, where R_{rt} is defined by [1, eqn. (23)].

For our purpose, we also consider a virtual system consisting of infinitely many PSCs. Among those PSCs, PSC l is a Rayleigh fading channel with perfect receive channel state information (CSI), and with independent and identically distributed (i.i.d.) phase-shift keying (PSK) inputs of average SNR $\rho[l]$ as defined by [1, Eqn. (19)]. Furthermore, the fading process within each PSC is memoryless, and the noise samples are also i.i.d. zero-mean circular complex Gaussian. By such a construction, the L-average capacity, namely the average input-constrained capacity of the first L PSCs in this virtual system approaches R_{rt}, as L grows asymptotically large (cf. [1, Sec. III.A]). In our proof, we hence fix a sufficiently large L such that the L-average capacity of the virtual system exceeds $(1 - \lambda/2) \cdot R_{\mathrm{rt}}$.

In order to complete the achievability proof, it suffices to show that PSC l in the recursive training scheme achieves at least the input-constrained capacity of memoryless PSC l in the virtual system, for every l from 0 to $(L - 1)$. In light of [1, Eqn. (18)-(19)] and the one-step MMSE prediction in recursive training, we can write the channel equation of PSC l as

$$X[l; k] = \sqrt{\rho[l]} \cdot \hat{H}_{\mathrm{d}}[l; k] \cdot S[l; k] + \bar{Z}[l; k], \quad k = 1, \ldots, K, \qquad (1)$$

where the index $[l; k]$ denotes the kth channel use of PSC l. Due to the wide-sense stationarity and ergodicity of the original fading process, both $\{\hat{H}_{\mathrm{d}}[l; :]\}$ and $\{\bar{Z}[l; :]\}$ sequences are wide-sense stationary and ergodic with respect to index k. In this addendum, we further normalize them to be of unit variance and thus $\rho[l]$ is the average SNR. However, due to the non-negligible residual correlation among channel uses within PSC l, in general neither $\{\hat{H}_{\mathrm{d}}[l; :]\}$ nor $\{\bar{Z}[l; :]\}$ sequence can be a circular complex Gaussian process. Fortunately, we have and shall utilize a "locally" Gaussian property of (1). That is, for each k, $\hat{H}_{\mathrm{d}}[l; k]$ and $\bar{Z}[l; k]$ are jointly Gaussian and in fact independent, as shown in [1, Sec. III.A].

Let us evaluate the GMI of (1) following [3, Sec. III]. For a given codebook $\{s_m[l; :]\}_{m=1}^{M}$ and a given realization of received channel outputs $x[l; :]$ along with channel fading process $\hat{h}_{\mathrm{d}}[l; :]$, a mismatched channel decoder

3

chooses the message m that minimizes the metric

$$D(m) = \frac{1}{K} \sum_{k=1}^{K} \left| x[l;k] - \sqrt{\rho[l]} \cdot \hat{h}_{\mathrm{d}}[l;k] \cdot s_m[l;k] \right|^2. \tag{2}$$

That is, the decoder treats the PSC as if it is were memoryless Rayleigh fading with i.i.d. circular complex Gaussian noise. Consider the ensemble of all codebooks i.i.d. generated by a prescribed PSK constellation, and without loss of generality assume the first message is selected with its corresponding codeword transmitted. First, due to ergodicity, the metric $D(1)$ converges almost surely to unity, since

$$
\begin{aligned}
T &= \lim_{K \to \infty} D(1) \\
&= \lim_{K \to \infty} \frac{1}{K} \sum_{k=1}^{K} \left| \bar{Z}[l;k] \right|^2 \\
&= \mathbf{E} \left\{ \left| \bar{Z}[l;k] \right|^2 \right\} = 1, \quad \text{a.s.} \tag{3}
\end{aligned}
$$

The probability that an incorrect codeword accumulates a metric smaller than unity decays exponentially in K, and the GMI is just the exponent given by

$$I_{\mathrm{GMI}} = \sup_{\mu < 0} (\mu - \Lambda(\mu)). \tag{4}$$

The limiting log-moment generating function $\Lambda(\mu) = \lim_{K \to \infty} \frac{1}{K} \Lambda_K(K\mu)$ is induced via

$$\Lambda_K(\mu) = \log \mathbf{E} \left\{ e^{\mu D(m')} \Big| X[l;:], \hat{H}_{\mathrm{d}}[l;:] \right\}, \tag{5}$$

for $m' > 1$. Because the inputs within each codeword are i.i.d. PSK symbols, conditioning upon $X[l;:]$ and $\hat{H}_{\mathrm{d}}[l;:]$, we have

$$\frac{1}{K} \Lambda_K(K\mu) = \frac{1}{K} \sum_{k=1}^{K} \log \mathbf{E} \left\{ e^{\mu \left| X[l;k] - \sqrt{\rho[l]} \cdot \hat{H}_{\mathrm{d}}[l;k] \cdot S \right|^2} \Big| X[l;k], \hat{H}_{\mathrm{d}}[l;k] \right\}, \tag{6}$$

where the expectation is with respect to the PSK constellation S. For concreteness, let us denote the PSK constellation by $S \in \{\theta_j\}_{j=1}^{J}$ with each constellation point selected equiprobably. Hence we can further write (6) as

$$\frac{1}{K} \Lambda_K(K\mu) = \frac{1}{K} \sum_{k=1}^{K} \log \sum_{j=1}^{J} e^{\mu \left| X[l;k] - \sqrt{\rho[l]} \cdot \theta_j \cdot \hat{H}_{\mathrm{d}}[l;k] \right|^2} - \log J. \tag{7}$$

4

Due to ergodicity, for every $\mu < 0$ within a certain finite neighborhood of zero, $\Lambda(\mu)$ converges almost surely to an expectation, as

$$
\begin{aligned}
\Lambda(\mu) &= \lim_{K \to \infty} \frac{1}{K} \Lambda_K(K\mu) \\
&= \mathbf{E} \left\{ \log \sum_{j=1}^{J} e^{\mu \left| X - \sqrt{\rho[l]} \cdot \theta_j \cdot \hat{H}_{\mathrm{d}} \right|^2} \right\} - \log J, \quad \text{a.s.}, \quad (8)
\end{aligned}
$$

where the expectation is with respect to the fading coefficient $\hat{H}_{\mathrm{d}} \sim \mathcal{CN}(0,1)$ and the channel output corresponding to the first message, $X = \sqrt{\rho[l]} \cdot \hat{H}_{\mathrm{d}} \cdot S_1 + \bar{Z}$ with $\bar{Z} \sim \mathcal{CN}(0,1)$ independent of $\hat{H}_{\mathrm{d}}$. Note that the probability density function of X conditioned upon $\hat{H}_{\mathrm{d}}$ is

$$
p_{X|\hat{H}_{\mathrm{d}}}(x|\hat{h}_{\mathrm{d}}) = \frac{1}{J} \sum_{j=1}^{J} \frac{1}{\pi} e^{-\left| x - \sqrt{\rho[l]} \cdot \theta_j \cdot \hat{h}_{\mathrm{d}} \right|^2}. \quad (9)
$$

So by letting $\mu = -1$, we have

$$
\begin{aligned}
I_{\mathrm{GMI}} &\geq -1 - \Lambda(-1) \\
&= -1 - \log \pi - \mathbf{E} \left\{ \mathbf{E} \left\{ \log p_{X|\hat{H}_{\mathrm{d}}}(X|\hat{H}_{\mathrm{d}}) \right\} \Big| \hat{H}_{\mathrm{d}} \right\} \\
&= h(X|\hat{H}_{\mathrm{d}}) - h(\bar{Z}) \\
&= h(X|\hat{H}_{\mathrm{d}}) - h(X|S_1, \hat{H}_{\mathrm{d}}) = I(S_1; X|\hat{H}_{\mathrm{d}}), \quad (10)
\end{aligned}
$$

which is precisely the input-constrained capacity of memoryless PSC l in the virtual system. Hence as we choose a sufficiently large coding blocklength K such that for every l from 0 to $(L-1)$, there exist a codebook and an associated mismatched channel decoder that achieve a fraction of $(1-\lambda)/(1-\lambda/2)$ of the GMI of PSC l with an average decoding error probability no greater than δ/L, the entire recursive training scheme achieves the rate $(1-\lambda) \cdot R_{\mathrm{rt}}$ with an average decoding error probability no greater than δ, following a standard union upper bounding argument. This concludes our proof.

References

[1] W. Zhang and J. N. Laneman, "How Good is PSK for Peak-Limited Fading Channels in the Low-SNR Regime?" *IEEE Trans. Inform. Theory*, Vol. 53, No. 1, pp. 236–251, Jan. 2007.

[2] R. G. Gallager, *Information Theory and Reliable Communication*, John Wiley & Sons, Inc., New York, NY, 1968.

[3] A. Lapidoth and S. Shamai (Shitz), "Fading Channels: How Perfect Need 'Perfect Side Information' Be?" *IEEE Trans. Inform. Theory*, Vol. 48, No. 5, pp. 1118–1134, May 2002.